ALLOSTERICALLY REGULATED RIBOZYMES

FIELD OF THE INVENTION

The present invention relates generally to the field of ribozymes and in particular to aptazymes or allosteric, regulatable ribozymes that modulate their kinetic parameters in response to the presence of an effector molecule.

BACKGROUND OF THE INVENTION

The United States Government may own certain rights in this invention under DARPA Grant No.: N65236-98-1-5413 and MURI Grant No.: DAAD19-99-1-0207.

This patent claims priority from Provisional Patent Application Serial No. 60/212,097, entitled "Aptazymes for Genetic Regulatory Circuits", filed June 15, 2000.

Ribozymes or RNA enzymes are oligonucleotides of RNA that can act like enzymes by catalyzing the cleavage of RNA molecules. Generally, ribozymes have the ability to behave like an endoribonuleases. The location of the cleavage site is highly sequence specific, approaching the sequence specificity of DNA restriction endonucleases. By varying conditions, ribozymes can also act as polymerases or dephosphorylases.

Ribozymes were first described in connection with *Tetrahymena thermophilia*. The Tetrahymena rRNA was shown to contain an intervening sequence (IVS) capable of excising itself out of a large ribosomal RNA precursor. The IVS is a catalytic RNA molecule that mediates self-splicing out of a precursor, whereupon it converts itself into a circular form. The Tetrahymena IVS is more commonly known now as the Group I Intron.

Regulatable ribozymes have been described, wherein the activity of the ribozyme is regulated by a ligand binding moiety. Upon binding the ligand, the ribozyme activity on a target RNA is changed. The ligand-binding portion is an RNA sequence capable of binding a ligand such as an organic or inorganic molecule, or even a prodrug. The regulatable ribozymes described to date target bind, e.g., a first target sequence and the enzymatic activity is brought to bear on a separate RNA molecule for cleavage.

SUMMARY OF THE INVENTION

The present invention provides an allosterically regulatable ribozyme or aptazyme that has the advantage of specific gene recognition with modulation of the enzymatic activity of the gene product typically exploited by pharmaceuticals. The aptazymes of the present invention are, therefore, allosteric ribozymes in that their activity is under the allosteric control of a second portion of the ribozyme. Just as allosteric protein enzymes undergo a change in their kinetic parameters or of their enzymatic activity in response to interactions with an effector molecule, the catalytic abilities of the regulatable aptazymes may similarly be modulated by an allosteric effector(s). Thus, the present invention is directed to allosterically regulatable aptazymes that transduce molecular recognition into catalysis.

The present invention includes an allosterically regulatable aptazyme construct that is inserted into a gene of interest, *e.g.*, a gene targeting expression vector. The regulatable aptazyme sequence provides gene specific recognition as well as modulation of the aptazyme's kinetic parameters. The kinetic parameters of the regulatable aptazyme vary in response to an allosteric effector molecule. Specifically, in the presence of the allosteric effector, the

aptazyme splices itself out of the gene in response to the effector molecule to regulate expression of the gene. An important feature of the present invention is that the regulatable aptazymes disclosed herein only require recognition rather than actual binding.

A key distinction with known systems is that the regulatory domains of the regulatable aptazymes of the present invention may bind targets, but they are engineered and selected without the necessity of knowing anything about their binding target. In the present invention it is the allosteric interaction of an effector molecule with the regulatory domain that transduces interactions into catalysis. Therefore, binding is a concomitant but secondary function of such interactions; that is, the allosteric ribozymes disclosed herein may bind the effector or the target very poorly, but upon their interaction, a synergistic effect is found that could not be detected by screening for each characteristic alone.

In the present invention the effector molecule does not produce a conformational change, but rather will add essential catalytic sites (e.g., residues) for a reaction. That is, both the effector molecule and the regulatable aptazyme contribute a portion of the active site of the ribozyme. For example, using the method of the present invention a ribozyme

and an effector molecule that would be unable to bind and/or perform an enzymatic function independently, may be isolated that act synergistically.

The present invention includes an aptazyme construct with a regulatable aptamer oligonucleotide sequence having a regulatory domain, such that the kinetic parameters of the aptazyme on a target gene vary in response to the interaction of an allosteric effector molecule with the regulatory domain.

The construct of one embodiment of the present invention provides a DNA oligonucleotide coding for an aptazyme domain so that the DNA can be transcribed to RNA (e.g., mRNA), where the RNA contains a self-splicing aptazyme domain that can be activated in the presence of an effector molecule.

Aptazymes are more robust than allosteric protein enzymes in several ways: (1) they can be selected in vitro, which facilitates the engineering of particular constructs; (2) the levels of catalytic modulation are much greater for aptazymes than for protein enzymes; and (3) since aptazymes are nucleic acids, they can potentially interact with the genetic machinery in ways that protein molecules may not.

At least part of the utility of the present invention is for use in the identification, isolation and enhancement of allosteric effectors and of the allosterically regulatable

aptazymes with which they interact. Similarly, it is possible to activate or repress a reporter gene (e.g., luciferase) containing an engineered intron in response to an endogenous activator. In this way, luciferase-engineered intron constructs may be used to monitor intracellular levels of proteins or small molecules such as cyclic AMP. Such applications may be used for high-throughput screening.

The present invention includes a regulatable aptazyme construct having an aptamer oligonucleotide sequence with a regulatory domain. A characteristic of the regulatable aptazyme construct of the present invention is that the kinetic parameters of the aptazyme vary in response to an effector molecule. In particular, the kinetic parameters of the aptazyme on a target gene vary in response to the interaction of an allosteric effector molecule with the regulatory domain. For example, the aptazyme splices itself out of a gene in response to the effector molecule interacting with the regulatory domain of the aptazyme to regulate expression of the target gene.

The present invention also contemplates that the regulatable aptazyme construct may be amplified by polymerase chain reaction. Finally, the invention contemplates that the regulatable aptamer oligonucleotide sequence of the construct

may include RNA nucleotides, so that the invention further includes reverse transcription of the RNA using reverse transcriptase to produce an DNA aptamer complementary to the RNA template.

Aptazymes, or allosterically activated ribozymes, have been developed that are activated by cyclic nucleotide monophosphates as well as other small molecules like theophylline. In addition to aptazymes activated by small molecules, there are natural ribozymes that are extremely dependent on proteins for their activity.

Other methods for aptazyme development using small molecule ligands have proven successful. In particular, it has been possible to add aptamer moieties to ribozymes, without selection, and achieve activation in the presence of ligand (like ATP or theophylline). One of the unique features of the present selection protocol relative to others that have previously been published is that the present invention randomizes not only a domain that is pendant on the catalytic core, but a portion of the catalytic core itself.

It should be noted that the method is not limited to RNA pools, but could also encompass DNA pools or modified RNA pools. The method is not limited to ligases, but could also encompass other ribozyme classes.

Nucleic acids are generally less robust than enzymes. However, modified nucleotides may be introduced into the aptazymes that substantially stabilize them from degradation in environments such as sera or urine. Similarly, enzymes generally have higher affinities for analytes than do aptamers, and be inference aptazymes. However, the analytical methods of the present invention do not rely on binding per se, but only on transient interactions. The present invention requires mere recognition rather that actual binding, providing a potential advantage of apatzymes over enzymes.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the features and advantages of the present invention, reference is now made to the detailed description of the invention along with the accompanying figures in which corresponding numerals in different figures refer to corresponding parts and in which:

Figure 1 is a depiction of the secondary structure of the Group 1 theophylline-dependent (td) intron of bacteriophage T4 (wild type);

Figure 2a is a photograph of a gel showing activation of the GpITh1P6.131 aptamer construct, together with a graphical representation of the gel, of one embodiment of the present invention;

Figure 2b is a photograph of a gel showing activation of GpITh2P6.133 aptamer construct, together with a graphical representation of the gel of one embodiment of the present invention.

Figure 3a depicts a portion of the P6 region of the Group I ribozyme joined to the anti-theophylline aptamer by a short randomized region to generate a pool of aptazymes of the present invention.

Figure 3b is a schematic depiction of a selection protocol for the Group I P6 Aptazyme Pool of Fig. 3a.

Figure 4 is a diagram of one embodiment of the present invention depicting exogenous or endogenous activation of Group I intron splicing;

Figure 5 is a diagram of another embodiment of the present invention depicting a strategy for screening libraries of exogenous activators; and

Figure 6 is a diagram of an alternative embodiment of the present invention for screening libraries of exogenous activators.

DETAILED DESCRIPTION OF THE INVENTION

While the making and using of various embodiments of the present invention are discussed in detail below, it should be appreciated that the present invention provides many applicable inventive concepts that may be embodied in a wide variety of specific contexts. The specific embodiments discussed herein are merely illustrative of specific ways to make and use the invention and do not delimit the scope of the invention.

Definitions

As used herein, the term "regulatable aptazyme" means an allosteric ribozyme. The kinetic parameters of the ribozyme may be varied in response to the amount of an allosteric effector molecule. Just as allosteric protein enzymes undergo a change in their kinetic parameters or of their enzymatic activity in response to interactions with an effector molecule, the catalytic abilities of regulatable aptazymes can similarly be modulated by effectors. Regulatable aptazymes transduce molecular recognition into catalysis upon interaction with an allosteric effector molecule that binds an effector portion of the regulatable aptazyme. Specifically, in the presence of the effector, the aptazyme splices itself

out of a gene in response to the effector molecule to regulate expression of the gene.

As used herein, the term "aptamer" refers to an oligonucleotide having aptazyme activity.

As used herein, the term "allosteric effector" or "allosteric effector molecule" means a substance that allosterically changes the kinetic parameters or catalytic activity of an aptazyme, and in particular a substance that activates self-splicing of an aptazyme.

As used herein, the term "kinetic parameters" refers to catalytic activity. Changes in the kinetic parameters of a catalytic ribozyme produce changes in the catalytic activity of the ribozyme such as a change in the rate of reaction or a change in substrate specificity. For example, self-splicing of an aptazyme out of a gene environment may result from a change in the kinetic parameters of the aptazyme.

As used herein, the term "catalytic" or "catalytic activity" refers to the ability of a substance to affect a change in itself or of a substrate under permissive conditions.

As used herein, the term "protein-protein complex" or "protein complex" refers to an association of more than one protein. The proteins of the complex may be associated by a

variety of means, or by any combination of means, including but not limited to functional, stereochemical, conformational, biochemical, or electrostatic association. It is intended that the term encompass associations of any number of proteins.

As used herein the terms "protein", "polypeptide" or "peptide" refer to compounds comprising amino acids joined via peptide bonds and are used interchangeably.

As used herein, the term "endogenous" refers to a substance the source of which is from within a cell. Endogenous substances are produced by the metabolic activity of a cell. Endogenous substances, however, may nevertheless be produced as a result of manipulation of cellular metabolism to, for example, make the cell express the gene encoding the substance.

As used herein, the term "exogenous" refers to a substance the source of which is external to a cell. An exogenous substance may nevertheless be internalized by a cell by any one of a variety of metabolic or induced means known to those skilled in the art.

As used herein, the term "gene" means the coding region of a deoxyribonucleotide sequence encoding the amino acid sequence of a protein. The term includes sequences located

adjacent to the coding region on both the 5', and 3', ends such that the deoxyribonucleotide sequence corresponds to the length of the full-length mRNA for the protein. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene which are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed, excised or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the MRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain

sequences which direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotides referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand.

The term "gene of interest" as used here refers to a gene, the function and/or expression of which is desired to be investigated, or the expression of which is desired to be regulated, by the present invention. In the present disclosure, the *td* gene of the T4 bacteriophage is an example

of a gene of interest and is described herein to illustrate the invention. The present invention may be useful in regard to any gene of any organism, whether of a prokaryotic or eukaryotic organism.

The term "hybridize" as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acid strands) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the melting temperature of the formed hybrid, and the G:C (or U:C for RNA) ratio within the nucleic acids.

The terms "complementary" or "complementarity" as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, for the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be partial, in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of

hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). A partially complementary sequence is one that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid; it is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous sequence or probe to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity

(*e.g.*, less than about 30% identity); in the absence of non-specific binding, the probe will not hybridize to the second non-complementary target sequence. When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described.

As known in the art, numerous equivalent conditions may be employed to comprise either low or high stringency conditions. Factors such as the length and nature (DNA, RNA, base composition) of the sequence, nature of the target (DNA, RNA, base composition, presence in solution or immobilization, etc.), and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate and/or polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. With "high

stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "weak" or "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less.

Low stringency conditions comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4 \cdot H_20$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent (50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA [Fraction V; Sigma]) and 100 µg/ml denatured salmon sperm DNA) followed by washing in a solution comprising 5X SSPE, 01% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

Numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to

generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions which promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

The term "antisense," as used herein, refers to nucleotide sequences that are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to tile "sense" strand. Antisense molecules may be produced by any method, including synthesis by ligating the gene(s) of interest in a reverse orientation to a viral promoter which permits the synthesis of a complementary strand. Once introduced into a cell, the transcribed strand combines with natural sequences produced by the cell to form duplexes. These duplexes then block either the further transcription or translation. In this manner, mutant phenotypes may also be generated. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term also is used in reference to RNA sequences that are complementary to a specific RNA sequence (*e.g.*, mRNA).

Included within this definition are antisense RNA ("asRNA") molecules involved in genetic regulation by bacteria. Antisense RNA may be produced by any method, including synthesis by splicing the gene(s) of interest in a reverse orientation to a viral promoter which permits the synthesis of a coding strand. Once introduced into an embryo, this transcribed strand combines with natural mRNA produced by the embryo to form duplexes. These duplexes then block either the further transcription of the mRNA or its translation. In this manner, mutant phenotypes may be generated. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. The designation. (-) (i.e., "negative") is sometimes used in reference to the antisense strand with the designation (+) sometimes used in reference to the sense (i.e., "positive") strand.

A gene may produce multiple RNA species which are generated by differential splicing of the primary RNA transcript. cDNAs that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cDNAs) and regions of complete non-identity (for example, representing the presence of exon "A" on cDNA I

wherein cDNA 2 contains exon "B" instead). Because the two cDNAs contain regions of sequence identity they will both hybridize to a probe derived from the entire gene or portions of the gene containing sequences found on both cDNAs; the two splice variants are therefore substantially homologous to such a probe and to each other.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the host cell being transformed and may include, but is not limited to, viral infection, electroporation, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. The term "transfection" as used herein refers to the introduction of foreign DNA into eukaryotic cells.

Transfection may be accomplished by a variety of means known to the art including, e.g., calcium phosphate-DNA co-

precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection, and biolistics. Thus, the term "stable transfection" or "stably transfected" refers to the introduction and integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell which has stably integrated foreign DNA into the genomic DNA. The term also encompasses cells which transiently express the inserted DNA or RNA for limited periods of time. Thus, the term "transient transfection" or "transiently transfected" refers to the introduction of foreign DNA into a cell where the foreign DNA fails to integrate into the genome of the transfected cell. The foreign DNA persists in the nucleus of the transfected cell for several days. During this time the foreign DNA is subject to the regulatory controls that govern the expression of endogenous genes in the chromosomes. The term "transient transfectant" refers to cells which have taken up foreign DNA but have failed to integrate this DNA.

As used herein, the term "selectable marker" refers to the use of a gene that encodes an enzymatic activity and which confers the ability to grow in medium lacking what would

otherwise be an essential nutrient (e.g., the HIS3 gene in yeast cells); in addition, a selectable marker may confer resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed. A review of the use of selectable markers in mammalian cell lines is provided in Sambrook, J. *et. a!., Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor Laboratory Press, New York (1989) pp.16.9-16.15.

As used herein, the term "reporter gene" refers to a gene that is expressed in a cell upon satisfaction of one or more contingencies and which, upon expression, confers a detectable phenotype to the cell to indicate that the contingencies for expression have been satisfied. For example, the gene for Luciferase confers a luminescent phenotype to a cell when the gene is expressed inside the cell. In the present invention, the gene for Luciferase may be used as a reporter gene such that the gene is only expressed upon the splicing out of an intron in response to an effector. Those cells in which the effector activates splicing of the intron will express Luciferase and will glow.

As used herein, the term "vector" is used in reference to nucleic acid molecules that transfer DNA segment(s) from one

cell to another. The term "vehicle" is sometimes used interchangeably with "vector."

The term "vector" as used herein also includes expression vectors in reference to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

As used herein, the term "amplify", when used in reference to nucleic acids refers to the production of a large number of copies of a nucleic acid sequence by any method known in the art. Amplification is a special case of nucleic acid replication involving template specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic 'd strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer may be single stranded for maximum efficiency in amplification but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded.

Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.* ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

As used herein, the term "target" when used in reference to the polymerase chain reaction, refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction. Thus, the "target" is sought to be sorted oat from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence.

As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture

containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence.

To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified".

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as DCTP or DATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

The invention is now described in detail with the use of the td intron from T4 bacteriophage (Fig. 1) for illustrative purposes. The description is not intended to limit the scope of the invention or the claims appended hereto.

Figure 1 depicts the secondary structure of the td intron from bacteriophage T4 (GenBank # M12742), wherein "td" means theophylline-dependent. The td intron was selected to illustrate the present invention because, among other things, mutational analysis has identified regions of this intron that can be engineered and modified. See Salvo, et al., Deletion-tolerance and trans-splicing of the bacteriophage T4 intron.

Analysis of the P6-L6a region. *J. Mol. Biol.* **211**, 537-549 (1990) and Salvo et al., The P2 element of the td intron is dispensable despite its normal role in splicing. *J. Mol. Biol.* **267**, 2845-2848 (1992). Thus, aptamer domains or pools may be engineered into the T4 intron.

An anti-theophylline aptamer has been described by R.D. Jenison, et al., High-resolution molecular discrimination by RNA. *Science* **263**, 1425-1429 (1994). In the present invention, the anti-theophylline aptamer was mounted in two locations in the td intron, shown by the shaded portions of Fig. 1. One location was at the termini of P1 and the other location was within P6. The P1 constructs may enable ligand-dependent conformational changes that alter the conformation or register of the U:G base pair which is critical for splicing. The P6 region was selected because mutational analysis indicated that deletion of the P6 stem destabilizes the intron.

Referring now to Figs 2a and 2b, in the present invention P6 constructs were made so that Group I splicing was activated by the presence of theophylline in the range of approximately 9 to 19 fold over constructs grown in the absence of theophylline, as described in the examples below:

The following examples illustrate the present invention in the *td* gene system of T4. For a full understanding of the examples, refer to Figures 2a and 2b. The examples are provide for illustrative purposes and do not limit the scope of the present invention or the scope of the appended claims.

Example 1: GpITh1P6

Engineering of Group I Aptazymes

The first example illustrates how to make an aptazyme construct and demonstrates self-splicing of the aptazyme out of a gene in response to an effector molecule.

Construction of a regulatable aptazyme

Oligos GpIWt3.129: 5'-TAA TCT TAC CCC GGA ATT ATA TCC AGC TGC ATG TCA CCA TGC AGA GCA GAC TAT ATC TCC AAC TTG TTA AAG CAA GTT GTC TAT CGT TTC GAG TCA CTT GAC CCT ACT CCC CAA AGG GAT AGT CGT TAG-3' (SEQ ID NO: 1) and GpITh1P6.131: 5'-GCC TGA GTA TAA GGT GAC TTA TAC TTG TAA TCT ATC TAA ACG GGG AAC CTC TCT AGT AGA CAA TCC CGT GCT AAA TTA TAC CAG CAT CGT CTT GAT GCC CTT GGC AGA TAA ATG CCT AAC GAC TAT CCC TT-3' (SEQ ID NO: 2) were annealed and extended in a 30 μl reaction containing 100 pmoles of each oligo, 250 mM Tris-HCl (pH 8.3), 40 mM $MgCl_2$, 250 mM NaCl, 5 mM DTT, 0.2 mM each dNTP, 45 units of AMV reverse transcriptase (RT: Amersham Pharmacia Biotech,

Inc., Piscataway, NJ) at 37° C for 30 minutes. The extension reaction was diluted 1 to 50 in H_2O.

A PCR reaction containing 1 μl of the extension dilution, 500 mM KCl, 100 mM Tris-HCl, (pH 9.0), 1% Triton® x-100, 15 mM $MgCl_2$, 0.4 μM of GpIWt1.75: 5'-GAT AAT ACG ACT CAC TAT AGG GAT CAA CGC TCA GTA GAT GTT TTC TTG GGT TAA TTG AGG CCT GAG TAT AAG GTG-3' (SEQ ID NO:3), 0.4 μM of GpIWt4.89: 5'-CTT AGC TAC AAT ATG AAC TAA CGT AGC ATA TGA CGC AAT ATT AAA CGG TAG CAT TAT GTT CAG ATA AGG TCG TTA ATC TTA CCC CGG AA-3' (SEQ ID NO:4), 0.2 mM each dNTP and 1.5 units of Taq polymerase (Promega, Madison, WI) was thermocycled 20 times under the following regime: 94° C for 30 seconds, 45° C for 30 seconds, 72° C for 1 minute. The PCR reaction was precipitated in the presence of 0.2 M NaCl and 2.5 volumes of ethanol and then quantitated by comparison with a molecular weight standard using agarose gel electrophoresis.

The aptazyme construct was transcribed in a 10 μl high yield transcription reaction (AmpliScribe from Epicentre, Madison, WI. The reaction contained 500 ng PCR product, 3.3 pmoles of P^{32} [α-32 P]UTP (3000 Camel), 1X AmpliScribe transcription buffer, 10 mM DTT, 7.5 mM each NTP, and 1 μl AmpliScribe T7 polymerase mix. The transcription reaction was incubated at 37° C for 2 hours. One unit of RNase free-DNase

was added and the reaction returned to 37° C for 30 minutes. The transcription was then purified on a 6% denaturing polyacrylamide gel to separate the full length RNA from incomplete transcripts and spliced products, eluted and quantitated spectrophotometrically.

In vitro Assay

The RNA (4 pmoles/12 μl H_2O) was heated to 94° C for 1 minute then cooled to 37° C over 2 minutes in a thermocycler. The RNA was divided into 2 splicing reactions (9 μl each) containing 100 mM Tris-HCl (pH 7.45), 500 mM KCl and 15 mM $MgCl_2$, plus or minus theophylline (2 mM). The reactions were immediately placed on ice for 30 minutes. GTP (1 mM) was added to the reactions (final volume of 10 μl) and the reactions were incubated at 37° C for 2 hours.

The reactions were terminated by the addition of stop dye (10 μl) (95% formamide, 20 mM EDTA, 0.5% xylene cyanol, and 0.5% bromophenol blue). The reactions were heated to 70° C for 3 minutes and 10 μl was electrophoresed on a 6% denaturing polyacrylamide gel. The gel was dried, exposed to a phosphor screen and analyzed using a Molecular Dynamics Phosphorimager (Sunnyvale, CA).

Activation was determined from the amount of circular intron in each reaction. Circularized introns migrate slower

than linear RNA and can be seen as the bands above the dark bands of linear RNA in the +Theo lanes of the gels of Figs. 2a and. 2b.

Example 2: GpIP6Thpool

In vitro Selection of Group I Aptazymes

Example 2 illustrates how to generate a population of aptazymes so that there is variation in the nucleotide sequence of the aptamers. This example also illustrates how to select for phenotypes that are responsive to an effector molecule from among that population of aptazymes.

Construction of Pool

The construction of the pool was similar to the construction of the individual engineered aptazyme constructs. Oligos GpIWt3.129 and GpIThP6pool: 5'-GCC TGA GTA TAA GGT GAC TTA TAC TAG TAA TCT ATC TAA ACG GGG AAC CTC TCT AGT AGA CAA TCC CGT GCT AAA TN(1-4)A TAC CAG CAT CGT CTT GAT GCC CTT GGC AGN(1-4) TAA ATG CCT AAC GAC TAT CCC TT-3' (SEQ ID NO:5) were extended in the same manner as above. The extension reaction was diluted and used as template for a PCR reaction. The PCR reaction was similar to the reaction described with the following exceptions: the volume was doubled and GpIWt4.89 was replaced with GpIMutG.101: 5'-CTT AGC TAC AAT ATG AAC TAA CGT

AGC ATA TGA CGC AAT ATT AAA CGG TAG TAT TAT GTT CAG ATA AGG TCG TTA ATC TTA CCC CGG AAT TCT ATC CAG CT-3' (SEQ ID NO:6) in which there is an G to A mutation at the terminal residue of the intron. The pool had a diversity of $1.16X10^5$ molecules. RNA was made as described above.

In vitro Negative Selection

The RNA (10 pmoles/70 μl H_2O) was heated to 94° C for 1 minute then cooled to 37° C over 2 minutes in a thermocycler. The splicing reaction (90 μl) contained 100 mM Tris-HCl (pH 7.45), 500 mM KCl and 15 mM $MgCl_2$. The reaction was immediately placed on ice for 30 minutes. GTP (1 mM) was added to the reaction (final volume of 100 μl) and the reaction was incubated at 37° C for 20 hours. The reaction was terminated by the addition 20 mM EDTA and precipitated in the presence of 0.2 M NaCl and 2.5 volumes of ethanol. The reaction was resuspended in 10 μl H_2O and 10 μl stop dye and heated to 70° C for 3 minutes and was electrophoresed on a 6% denaturing polyacrylamide gel with Century™ Marker ladder (Ambion, Austin, TX). The gel was exposed to a phosphor screen and analyzed. The unreacted RNA was isolated from the gel, precipitated and resuspended in 10 μl of H_2O.

Positive Selection

The RNA (5 μl of negative selection) was heated to 94° C for 1 minute then cooled to 37° C over 2 minutes in a thermocycler. The positive splicing reaction (45 μl) contained 100 mM Tris-HCl (pH 7.45), 500 mM KCl, 15 mM $MgCl_2$ and 1mM theophylline. The reaction was immediately placed on ice for 30 minutes. GTP (1 mM) was added to the reaction (final volume of 50 μl) and the reaction was incubated at 37° C for 1 hour. The reaction was terminated by the addition of stop dye, heated to 70° C for 3 minutes and was electrophoresed on a 6% denaturing polyacrylamide gel with Century™ Marker ladder. The gel was exposed to a phosphor screen and analyzed. The band corresponding to the linear intron was isolated from the gel and precipitated and resuspended in 20 μl H_2O.

Amplification and Transcription

The RNA was reverse transcribed in a reaction containing 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM, $MgCl_2$, 0.1 M DTT, 0.4 mM of each dNTP 2 μM GpIMutG.101 and 400 units of SuperScript II reverse transcriptase (Gibco BRL, Rockville, MD). The cDNA was then PCR amplified, transcribed and gel purified as described above.

Figure 3a depicts the critical residues of the P6 region of the Group I ribozyme joined to the anti-theophylline aptamer by a short randomized region to generate a pool of aptazymes of the present invention. The residues shown in bold in Fig. 4a are the P6 critical residues, and the faded residues shown in Fig. 4a are the anti-theophylline aptamer. The randomized regions are designated in Fig. 4a as "N1-4". Approximately 40 random sequence residues are introduced into the N1-4 region of the construct to synthesize a pool of aptazymes, referred to herein as a communication module pool.

Figure 3b shows a selection protocol for the Group I P6 Aptazyme Pool of Fig. 3a. Positive and negative selections are made in vitro to select Group I aptazymes that are dependent on theophylline. The selections are described above in Example 2 for a specific embodiment of the present invention.

A communication module pool can be transformed with the selected aptazymes. The best theophylline-dependent Group I aptazymes that have been derived by any of the methods described herein may undergo further selection by partially randomizing their sequences and selecting for improved performance.

Strategies similar to those depicted in Figs 3a and 3b may be used to develop aptazymes and aptamers dependent on any desired effector molecule. See generally G.A. Soukup, et al., Engineering precision RNA molecular switches. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 3584-3589 (1999) and M. Koizumi, et al., Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nature Struct. Biol.* **6,** 1062-1071 (1999). Positive and negative in vitro selection such as depicted in Fig. 3b are described above in Example 2 for a specific embodiment of the present invention. The optimization strategies described herein yield Group I aptazymes that are highly dependent on small molecule effectors.

Figure 4 is a diagrammatic representation of one embodiment of the exogenous or endogenous activation of Group I intron splicing is depicted. A gene of interest 10 is fused to a reporter gene 12 such as luciferase or beta-galactosidase, which also contains the group I intron (td) 14. Splicing-out of the Group I intron is induced by an endogenous effector molecule 16. Alternatively, splicing-out of the Group I intron may be induced by an exogenous effector molecule 18. Activation of the aptazyme and auto-excision of the intron results in expression of the reporter gene encoded

protein 20 that is detect by, e.g., fluorescence 22 or any other desired detectable reaction.

Figure 5 is a diagram of another embodiment of the present invention. Libraries of candidate exogenous activators 30 can be generated from a randomized aptazyme pool indicated by random loop E_{1-n}. As in the embodiment of Fig. 4, a reporter gene 12 is expressed in cells where the exogenous activator 30 activates the aptazyme to release the intron, which may contain a random loop 32, from the gene. In this embodiment, the reaction occurs within cells which are then sorted 34 based on a chromogenic reaction or emission 22, or may even be isolated by, e.g, statistical cell separation cloning. As will be known to those of skill in the art of enzymatic oligonucleotides any number of current and future effector molecule libraries may be used with the present invention.

Figure 6 depicts an alternative embodiment for screening libraries of exogenous activators. Group I introns with length polymorphisms are induced into the construct by trans-splicing with an independent oligonucleotide. Libraries of candidate exogenous activators 30 can be generated from a randomized aptazyme pool indicated by random loop E_{1-n}. As in the embodiment of Fig. 5, a reporter gene 12 is expressed in

cells where the exogenous activator 30 activates the aptazyme to release the intron, which may contain a random loop 32, from the gene. In this embodiment, the reaction occurs within the intron 14 and an independent oligonucleotide 36 by a trans-splicing reaction and extraction step 38. Extracted trans-spliced intron reporter gene constructs are then amplified by, e.g., polymerase chain reaction in step 40, followed by transformation of cells with the transpliced construct at step 42. Transformation of the transpliced construct may be performed by those of skill in the art with either a negative or positive selection scheme for identification of the trans-spliced gene.

All publications mentioned in the above specification are hereby incorporated by reference. Modifications and variations of the described compositions and methods of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described compositions and modes of carrying out the invention which are obvious to those skilled in molecular biology or

related arts are intended to be within the scope of the following claims.